Going Deeper: Establishing Rules for the Ethical Use of Artificial Intelligence in Humanitarian Emergencies

Tino Kreutzer | @tinokreutzer

YORK UNIVERSITÉ UNIVERSITY

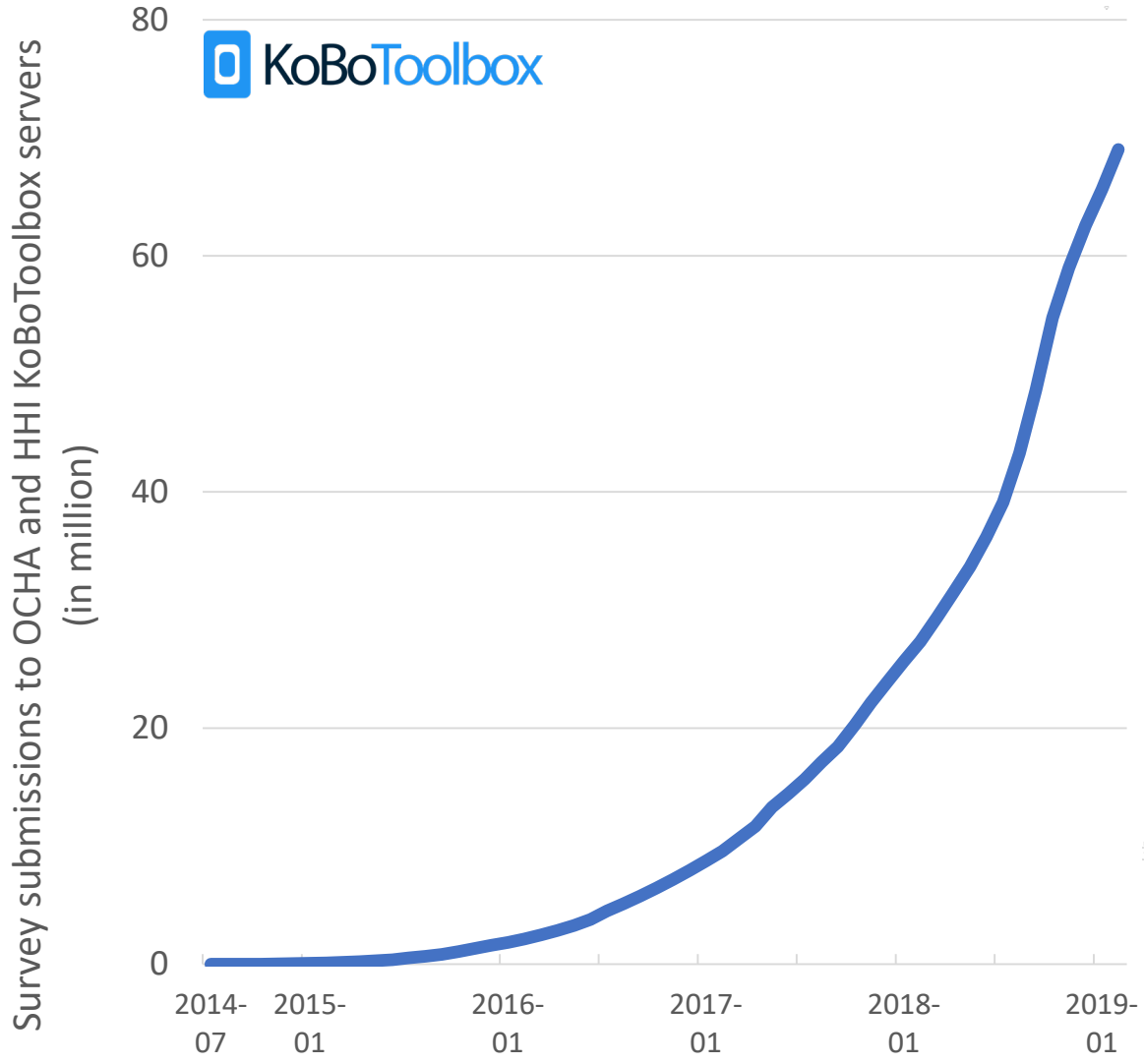
HARVARD HUMANITARIAN INITIATIVE

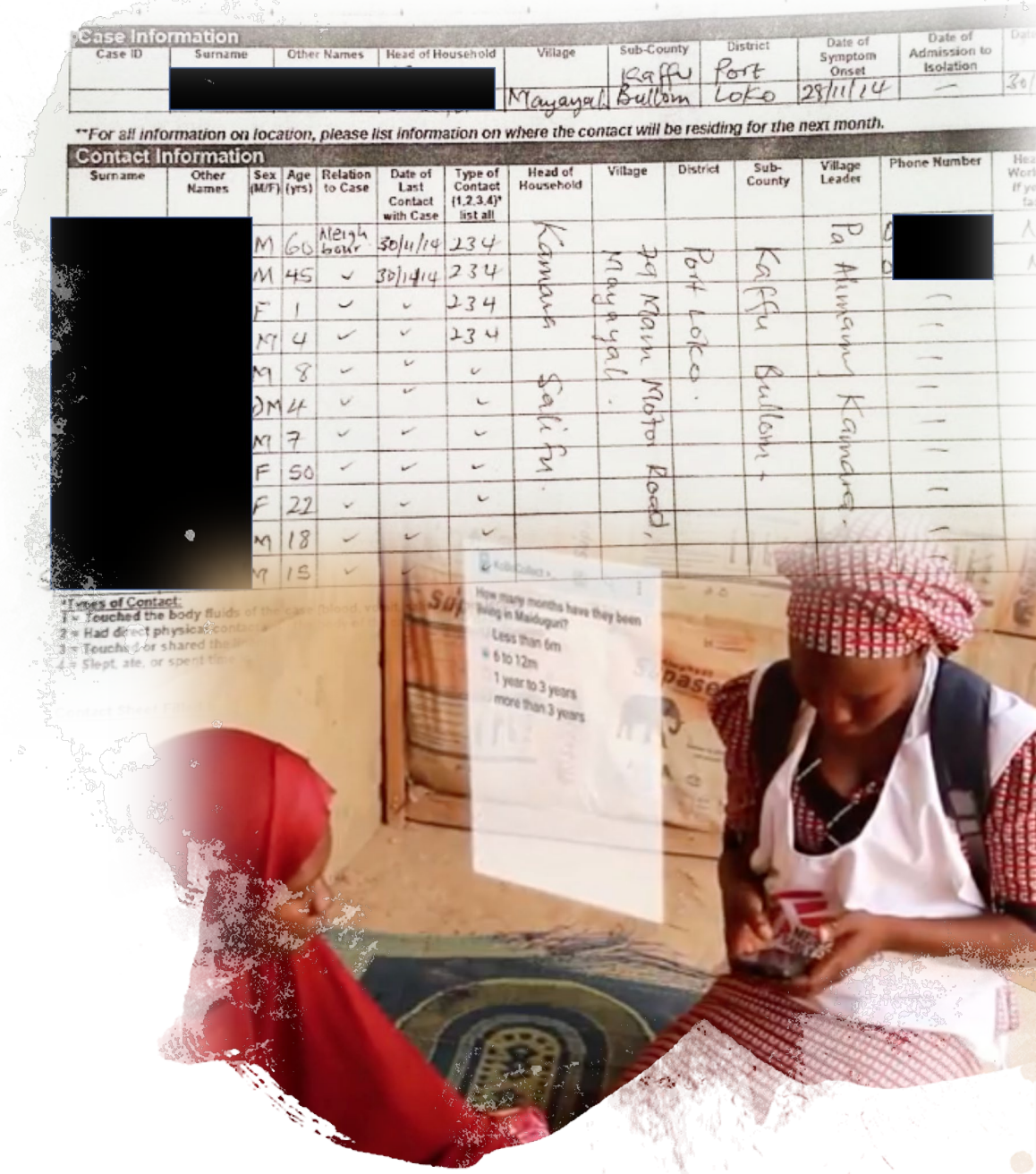
NETHOPE

This presentation

1. Techniques and technologies used for understanding affected populations in humanitarian emergencies
2. Natural Language Processing for qualitative data
3. Humanitarian and AI principles for ethical use of data and technology

From paper to mobile data collection to call centers






Beira, March 22, 2019

Maiduguri, 2017


Greece, 2016

M.9 How long does it take to collect water from main water source?

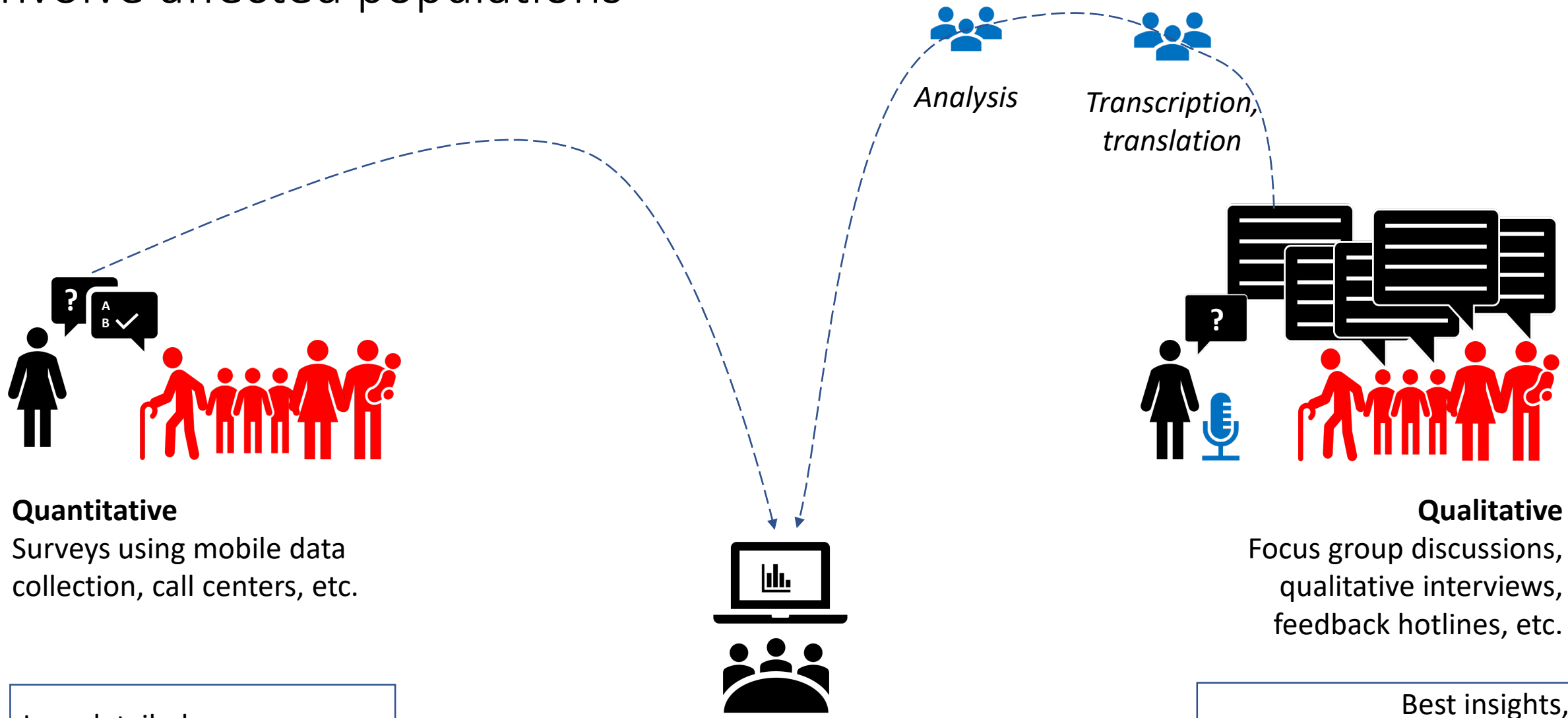
- (X) Less than 30 minutes
- ( ) 30 min-2 hours
- ( ) 2-4hours
- ( ) more than 4 hours
- ( ) DK

In the morning it can take a long time because I usually go to the well close to our house. But then the person who has the key to the lock may not be there so I need to wait a long time. Then I may need to come back in the afternoon. So quite often we go to the river which even though it's very far and we know that water can make us sick. Our ward chief has a water pump and the quality is very good, but he charges us a lot.

How long is the walk to the chief?

Oh, not far, only 10 minutes.

Primary data collection to understand and involve affected populations



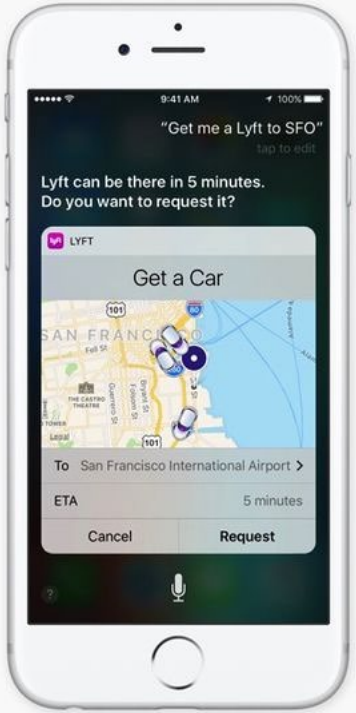
Quantitative
Surveys using mobile data collection, call centers, etc.

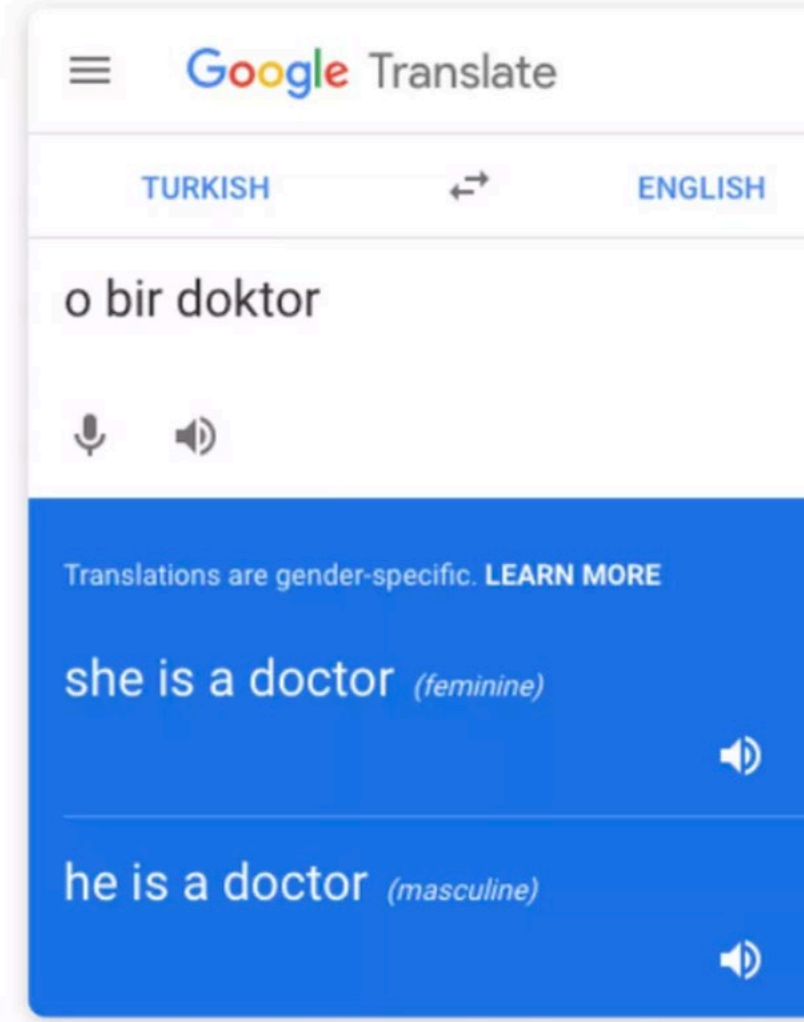
Less detailed,
but fast and easy to scale

Qualitative
Focus group discussions, qualitative interviews, feedback hotlines, etc.

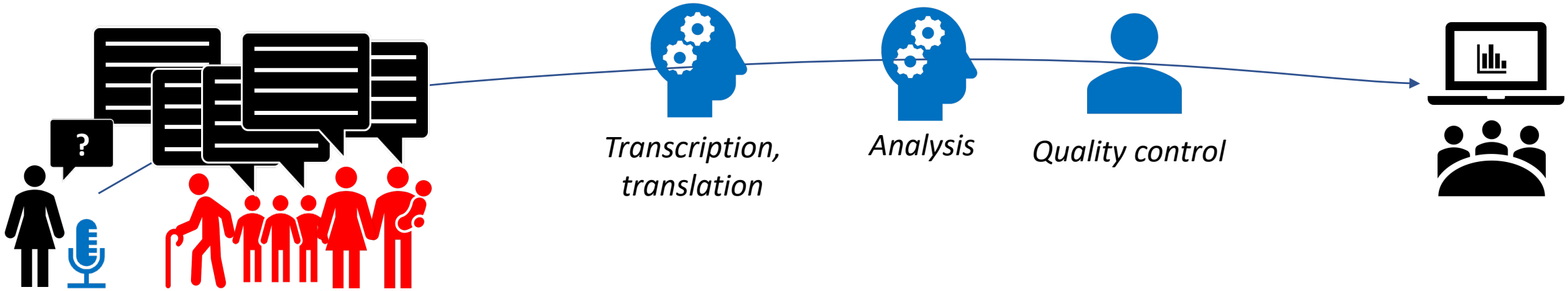
Best insights,
but hard to scale due to cost and lack of staff

Natural Language Processing





Understanding affected populations with NLP



Quantitative & Qualitative
Population surveys, key informant interviews, focus group discussions, qualitative interviews, feedback hotlines, etc.

NLP example for response analysis

water_source_primary
well

On a typical day, where do you usually get your water from, and are there other alternatives? Are there any issues with these sources?

water_source_distance
close

water_sources
well
river
pump

water_issues
long wait time
long distance
access fees
bad water quality

In the morning it can take a long time because I usually go to the well close to our house. But then the person who has the key to the lock may not be there so I need to wait a long time. Then I may need to come back in the afternoon. So quite often we go to the river which even though it's very far and we know that water can make us sick. Our ward chief has a water pump and the quality is very good, but he charges us a lot.

Moré tuáñr nam án koíba né?

Sorry, I didn’t get that.

Ethical risks of deploying new technologies in humanitarian response

- False sense of accuracy
- Increased remote management and ‘bunkerization’
- Risks to respondents and field staff due to collected data
- Testing and experimentation with the most vulnerable populations
- Exclusion of people without access to cell phones
- Exploitative data mining
- Technology can’t solve real social problems
- Uncritical involvement of large technology companies

“intelligent machines can restrict the choices of individuals and groups, lower living standards, disrupt the organization of labor and the job market, influence politics, clash with fundamental rights, exacerbate social and economic inequalities, and affect ecosystems, the climate and the environment”

(Montreal Declaration 2018)

What could go wrong with bringing AI to humanitarian emergencies???

Replacing biased humans with biased AI

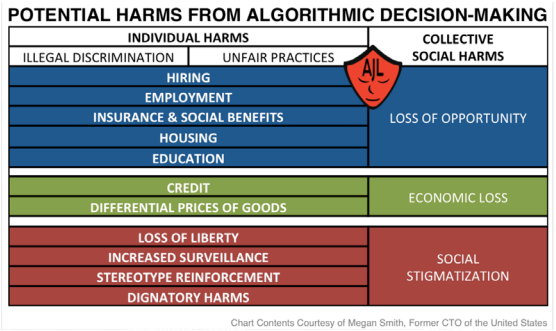
CNN BUSINESS Markets Tech Media Success Perspectives Video

Uber self-driving car kills pedestrian in first fatal autonomous crash

by Matt McFarland @mattmcfarland

March 19, 2018: 1:40 PM ET

Pedestrian killed in accident involving self-driving Uber



Gender Classifier	Darker Male	Darker Female	Lighter Male	Lighter Female	Largest Gap
Microsoft	94.0%	79.2%	100%	98.3%	20.8%
FACE++	99.3%	65.5%	99.2%	94.0%	33.8%
IBM	88.0%	65.3%	99.7%	92.9%	34.4%

Buolamwini 2018

Cornell University

the Simo

arXiv.org > cs > arXiv:1602.03506

Search or Artic

(Help | Advanced s

Computer Science > Artificial Intelligence

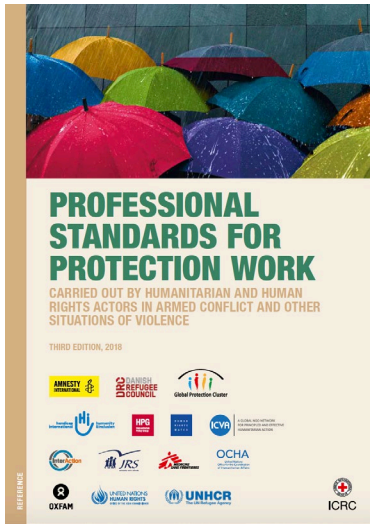
Research Priorities for Robust and Beneficial Artificial Intelligence

Stuart Russell (Berkeley), Daniel Dewey (FHI), Max Tegmark (MIT)

(Submitted on 10 Feb 2016)

Success in the quest for artificial intelligence has the potential to bring unprecedented benefits to humanity, and it is therefore worthwhile to investigate how to maximize these benefits while avoiding potential pitfalls. This article gives numerous examples (which should by no means be construed as an exhaustive list) of such worthwhile research aimed at ensuring that AI remains robust and beneficial.

Comments:	This article gives examples of the type of research advocated by the open letter for robust & beneficial AI at this http URL
Subjects:	**Artificial Intelligence (cs.AI)**; Machine Learning (stat.ML)
Journal reference:	AI Magazine 36:4 (2015)
Cite as:	**arXiv:1602.03506 [cs.AI]**
	(or **arXiv:1602.03506v1 [cs.AI]** for this version)



Data Protection & Humanitarian Principles

- IHL, Human Rights Law
- 1965 Humanitarian Principles: humanity, impartiality, neutrality, and independence
- 1994 Code of Conduct for the International Red Cross and Red Crescent Movement and NGOs in Disaster Relief
- **1998 Humanitarian Charter**
  - "people are not put at risk as a result of the way that humanitarian actors record and share information"
- **Protection Principles**
  - "minimise any negative effects of humanitarian action," such as rendering "civilians more vulnerable to attack"
- **2014 Core Humanitarian Standard**
  - requires organizations to "safeguard any personal information collected from communities and people affected by crisis that could put them at risk."
- **2014 Core Humanitarian Standard**
  - "clear and comprehensive policies on data protection" should be "aligned with international standards and local data protection laws."
- **2018 Professional Standards in Protection Work (ICRC)**

Data protection guidelines

Title	Year	Authors
Mapping and Comparing Responsible Data Approaches	2016	Berens, Mans, Verhulst
The Signal Code: A Human Rights Approach to Information During Crisis	2017	HHI / Signal
Professional Standards for Protection Work (3rd ed.)	2018	ICRC
ICRC Rules on Personal Data Protection	2017	ICRC
Handbook on Data Protection in Humanitarian Action	2017	ICRC
Data preparedness: connecting data, decision-making and humanitarian response	2016	Raymond and Al Achkar
Information and Data Management Training toolkit		Protection Cluster
Data Protection Starter Kit	2017	Terre des Hommes / CartONG
Responsible Data Principles	2018	The Engine Room
Big Data for Development and Humanitarian Action: Towards Responsible Governance	2016	UN Global Pulse
Policy on the Protection of Personal Data of Persons of Concern to UNHCR	2015	UNHCR
Conducting Mobile Surveys Responsibly: A Field Book for WFP Staff, May 2017	2017	WFP
Data Protection, Privacy and Security for Humanitarian & Development Programs	2017	WVI
Compromised Connections: Overcoming Privacy Challenges of the Mobile Internet	2016	Internews
The Humanitarian Metadata Problem - Doing No Harm in the Digital Era	2018	Privacy International and ICRC
ICT4Refugees: A report on the emerging landscape of digital responses to the refugee crisis	2016	Mason and Buchman
Technologies for monitoring in insecure environments	2016	Dette, Steets, Sagmeister
Building data responsibility into humanitarian action	2016	Raymond et al.

HANDBOOK ON DATA PROTECTION IN HUMANITARIAN ACTION

Co-editors: Christopher Kuner and Massimo Marelli

VUB BRUSSELS PRIVACY HUB

ICRC

REFERENCE

A Growing list of ethical AI guidelines

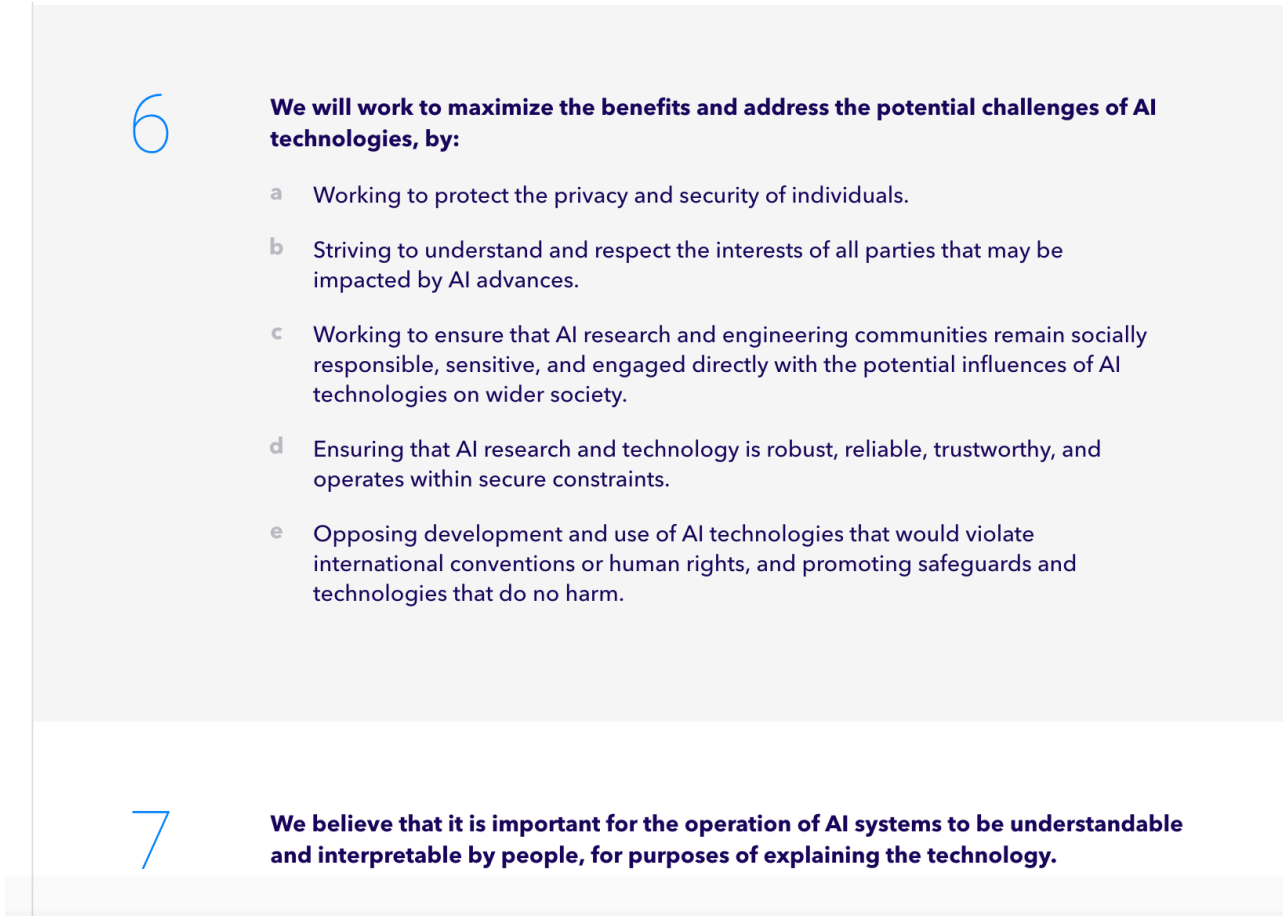
General principles

- 2017 *Asilomar Principles*
- 2018 Google *AI Principles*
- 2018 Microsoft: *Principles, Policies and Laws for the Responsible Use of AI*
- Partnership on AI: *Tenets*

Detailed guidance

- Association for Computing Machinery
- Institute of Electrical and Electronics Engineers (IEEE)
- Microsoft
- 2018 Montreal Declaration

Partnership on AI Tenets (examples)



6

We will work to maximize the benefits and address the potential challenges of AI technologies, by:

a Working to protect the privacy and security of individuals.

b Striving to understand and respect the interests of all parties that may be impacted by AI advances.

c Working to ensure that AI research and engineering communities remain socially responsible, sensitive, and engaged directly with the potential influences of AI technologies on wider society.

d Ensuring that AI research and technology is robust, reliable, trustworthy, and operates within secure constraints.

e Opposing development and use of AI technologies that would violate international conventions or human rights, and promoting safeguards and technologies that do no harm.

7

We believe that it is important for the operation of AI systems to be understandable and interpretable by people, for purposes of explaining the technology.

Asilomar Principles (extract)

Ethics and Values

6) **Safety:** AI systems should be safe and secure throughout their operational lifetime, and verifiably so where applicable and feasible.

7) **Failure Transparency:** If an AI system causes harm, it should be possible to ascertain why.

8) **Judicial Transparency:** Any involvement by an autonomous system in judicial decision-making should provide a satisfactory explanation auditable by a competent human authority.

9) **Responsibility:** Designers and builders of advanced AI systems are stakeholders in the moral implications of their use, misuse, and actions, with a responsibility and opportunity to shape those implications.

10) **Value Alignment:** Highly autonomous AI systems should be designed so that their goals and behaviors can be assured to align with human values throughout their operation.

11) **Human Values:** AI systems should be designed and operated so as to be compatible with ideals of human dignity, rights, freedoms, and cultural diversity.

12) **Personal Privacy:** People should have the right to access, manage and control the data they generate, given AI systems' power to analyze and utilize that data.

13) **Liberty and Privacy:** The application of AI to personal data must not unreasonably curtail people's real or perceived liberty.

14) **Shared Benefit:** AI technologies should benefit and empower as many people as possible.

15) **Shared Prosperity:** The economic prosperity created by AI should be shared broadly, to benefit all of humanity.

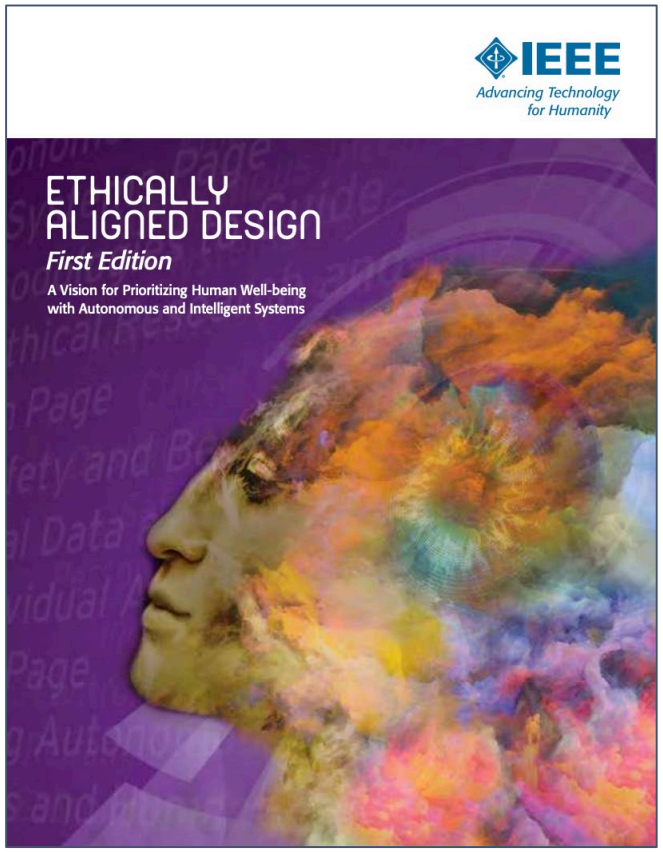
16) **Human Control:** Humans should choose how and whether to delegate decisions to AI systems, to accomplish human-chosen objectives.

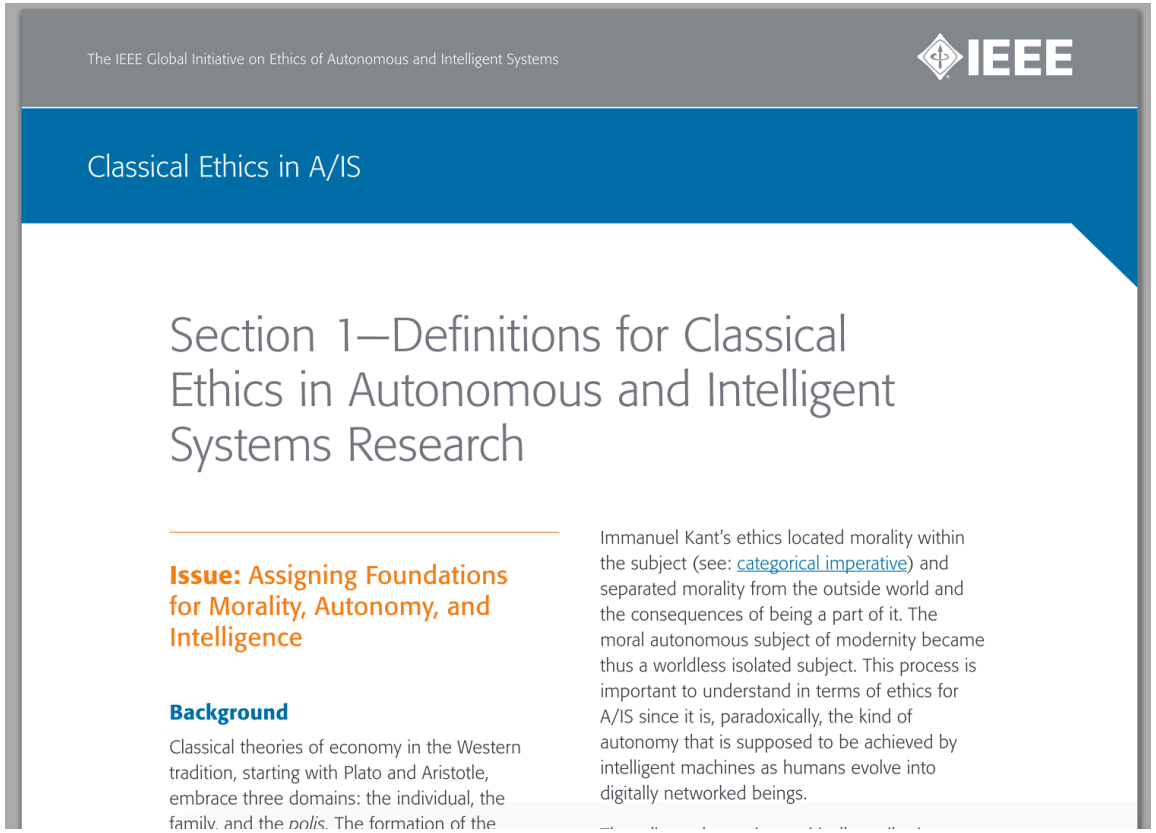
17) **Non-subversion:** The power conferred by control of highly advanced AI systems should respect and improve, rather than subvert, the social and civic processes on which the health of society depends.

18) **AI Arms Race:** An arms race in lethal autonomous weapons should be avoided.

IEEE Global Initiative on Ethics of Autonomous and Intelligent Systems

Goal: “provide pragmatic and directional insights and recommendations, serving as a key reference for the work of technologists, educators and policymakers in the coming years.”





Montreal AI declaration principles applicable to humanitarian response (examples)

3.1 Personal spaces in which people are not subjected to surveillance or digital evaluation must be protected from the intrusion of AIS and data acquisition and archiving systems (DAAS).

5.1 AIS processes that make decisions affecting a person's life, quality of life, or reputation must be intelligible to their creators.

5.2 The decisions made by AIS affecting a person's life, quality of life, or reputation should always be justifiable in a language that is understood by the people who use them or who are subjected to the consequences of their use. […]

6.1 AIS must be designed and trained so as not to create, reinforce, or reproduce discrimination based on — among other things — social, sexual, ethnic, cultural, or religious differences.

9.2 In all areas where a decision that affects a person's life, quality of life, or reputation must be made, where time and circumstance permit, the final decision must be taken by a human being and that decision should be free and informed

Proposed Humanitarian AI impact risk assessment framework

Impact level	Impact on affected people or communities	Example
I	Little to no impact; errors cause no harm	Transcription and classification of survey responses
II	Moderate impact; errors can cause minimal to moderate harm	Automatic analysis and routing of feedback/complaints received
III	High impact, errors can cause moderate to high harm	Automated decision-making that would decide whether an affected person or their household should receive particular assistance
IV	Very high impact; errors can cause serious to catastrophic harm	Donor-level decisions on attributing funding for particular sectors or emergencies

Humanitarian AI ethics principles

- AI is different from other “experimental” technologies. Risks are greater.
- Imagine all the ways how things could go wrong.
- Algorithmic decision-making in humanitarian response should require new ethical assessment framework
- There are many principles. Apply them. Teach them to your technology partners.
- Follow existing standards and guidance on data protection, data privacy, role of private vendors

Our road ahead…

Partnerships, research, ethical safeguards

Pilot

- Speech recognition for Yemeni Arabic
  - Manually transcribe and translate sample responses to serve as text training data
  - Collect speech recordings from volunteers as audio training data
- Response classification
  - Quality control for quantitative surveys
  - Question-specific analysis models
  - Validation and iterative improvements

Roll out

- Release of all language models, classification algorithms, and software API
- Provide documentation on how to create own classification models

Thank you!